

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
**FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO**

Endiel Silveira Spitznagel

**A importância da variação denominativa para recuperação  
da informação.**

PORTO ALEGRE  
2018

Endiel Silveira Spitznagel

## A importância da variação denominativa para recuperação da informação.

Trabalho apresentado como requisito parcial para obtenção do grau de bacharel em Biblioteconomia na Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

Orientadora: Dra. Rita do Carmo Ferreira  
Laipelt

Coorientadora: M.e Luciana Monteiro Krebs

PORTO ALEGRE  
2018

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

Reitor: Prof. Dr. Rui Vicente Oppermann

Vice-Reitora: Prof<sup>a</sup> Dr<sup>a</sup> Jane Fraga Tutikian

**FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO**

Diretora: Prof.<sup>a</sup> Dr<sup>a</sup>. Karla Maria Müller

Vice Diretora: Prof.<sup>a</sup> Dr<sup>a</sup>. Ilza Maria Tourinho Girardi

**DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO**

Chefe: Prof<sup>a</sup>. Dr<sup>a</sup>. Jeniffer Alves Cuty

Chefe Substituta: Prof<sup>a</sup>. Dr<sup>a</sup>. Eliane Lourdes da Silva Moro

**COMISSÃO DE GRADUAÇÃO DO CURSO DE BIBLIOTECONOMIA**

Coordenadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Rita do Carmo Ferreira Laipelt

Coordenador Substituto: Prof. Dr. Rene Faustino Gabriel Júnior

**Dados Internacionais de Catalogação na Publicação (CIP)**

Spitznagel, Endiel Silveira

A importância da variação denominativa para recuperação da informação / Endiel Silveira Spitznagel; Orientadora: Rita do Carmo Ferreira Laipelt; Coorientadora: Luciana Monteiro Krebs – Porto Alegre 2018.

51 f.

Monografia (Graduação) –  
Universidade Federal do Rio Grande do Sul, Faculdade  
de Biblioteconomia e Comunicação, Curso de  
Biblioteconomia, Porto Alegre, BR-RS, 2017.

1. Variação terminológica denominativa 2. Sistemas de recuperação da informação  
3. Ciência da informação 4. Terminologia I. Laipelt, Rita do Carmo Ferreira, orient. II. Título.

Endiel Silveira Spitznagel

## A importância da variação denominativa para recuperação da informação.

Trabalho apresentado como requisito parcial para obtenção do grau de bacharel em Biblioteconomia na Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

### BANCA EXAMINADORA

---

Profa. Dra. Rita do Carmo Ferreira Laipelt (Orientadora)  
Universidade Federal do Rio Grande do Sul

---

Prof. Me Luciana Monteiro Krebs (Coorientadora)  
Universidade Federal do Rio Grande do Sul

---

Prof. Dra. Sônia Elisa Caregnato  
Universidade Federal do Rio Grande do Sul

---

Prof. Dr. Rafael Port da Rocha

## AGRADECIMENTOS

Agradeço a minha orientadora Rita do Carmo Ferreira Laipel e a minha coorientadora Luciana Monteiro Krebs pela paciência e dedicação durante as orientações deste trabalho.

Agradeço também ao analista de sistema Vicente Grassi Filho que criou o software tornando para extração dos logs este trabalho possível.

Um agradecimento também aos meus pais que sempre me apoiaram.

## RESUMO

Estudo qualitativo que utiliza a análise de logs combinada com a metodologia de análise de variantes (Lapeilt, 2015) para identificar as variantes denominativas e suas tipologias e responder a questão problema: Que tipo de variantes terminológicas são utilizadas pelos usuários do portal LEXML? Tem como objetivo geral analisar as tipologias de variação denominativa da área do Direito Penal utilizadas pelos usuários do Portal LexMI e como objetivos específicos: identificar quais as tipologias de variantes denominativas na área do Direito Penal são utilizadas com maior frequência pelos usuários do Portal LexML e verificar se as tipologias de variantes identificadas ocorrem no Vocabulário Controlado do Senado Federal Brasileiro. Através da análise de logs do portal lexml foram encontrados como resultados alterações gráficas, reduções, Alterações léxicas de unidades poliléxicas e um caso excepcional que não se enquadra na classificação de Freixa (2002). Conclui que dos 16 termos nos 7 conjuntos 3 foram encontrados como autoridade, 3 como preferidos, 1 como variante e 9 não foram encontrados, demonstrando que as variantes denominativas afetam diretamente a recuperação da informação

Palavras-chave: Ciência da Informação. Sistemas de Recuperação da Informação. Terminologia. Variação denominativa. Direito Penal.

## Abstract

The importance of denominative variation to the information retrieval

A qualitative study using log analysis combined with the variance analysis methodology (Lapeilt, 2015) to identify the word variants and their typologies and answer the problem question: What kind of terminological variants are used by LEXML portal users? Its general objective is to analyze the typologies of denominational variation in the area of Criminal Law used by the users of the LexMI Portal and as specific objectives: To identify which typologies of word variants in the area of Criminal Law are used more frequently by users of the LexML Portal and verify if the typologies of variants identified occur in the Controlled Vocabulary of the Brazilian Federal Senate. Through analysis of lexml portal logs was found as results graphical changes, reductions, lexical changes of poly-xyx units and an exceptional case that does not fit the classification of Freixa (2002). It concludes that of the 16 terms in the 7 sets 3 were found as authority, 3 as preferred, 1 as a variant and 9 were not found, demonstrating that denominative variants directly affect information retrieval

Key words: information science. information retrieval systems. terminology. terminological denominative variation. criminal law.

## Sumário

1 INTRODUÇÃO.....	9
2 REFERENCIAL TEÓRICO.....	12
2.1 Sistemas de Recuperação da Informação .....	12
2.2 Indexação.....	14
2.3 Variação denominativa .....	19
2.4 Análise de Logs.....	27
2.4 Garantia do usuário .....	29
3 METODOLOGIA .....	33
3.1 Organização e Limpeza dos dados.....	34
3.2 Análise dos dados .....	36
4 RESULTADOS DA PESQUISA.....	38
4.1 Alterações Gráficas .....	38
4.2 Reduções .....	42
4.3 Alterações léxicas de unidades poliléxicas .....	43
4.4 Casos excepcionais.....	45
5 CONSIDERAÇÕES FINAIS .....	46
REFERÊNCIAS .....	49



## 1 INTRODUÇÃO

Ao longo do tempo a terminologia passou a contribuir com as Ciências da Informação, principalmente na área de recuperação da informação. A terminologia foi passando a fazer parte da construção de linguagens documentárias alfabéticas possibilitando uma melhoria na recuperação da informação.

Esta recuperação da informação ocorre em sistemas específicos que podem ou não ser de uma área do conhecimento, como é o caso do portal que foi utilizado para realização deste trabalho, o qual é especializado em direito.

Como será observado ao longo do trabalho, diversos fatores podem influenciar na recuperação da informação, a linguagem sendo um deles e um dos mais importantes. Dentro da linguagem podem ocorrer variações denominativas que é o uso de um termo diferente para expressar um mesmo conceito.

De acordo com Lancaster (1979, 2004), a indexação é o processo pelo qual a informação é inserida no sistema para posteriormente poder ser recuperada. Neste momento pode ser utilizada uma linguagem controlada, na forma de cabeçalhos de assunto ou tesauros.

A garantia do usuário é essencial na elaboração de linguagens controladas. É através dela que se consegue ter uma base das possíveis expressões de busca utilizadas, pois a utilização de termos muito genéricos ou muito específicos pode afetar o resultado das buscas. A decisão sobre qual utilizar depende da biblioteca em que o sistema está inserido.

Com o advento das tecnologias da informação foi se tornando cada vez mais necessário encontrar um meio de levantar os termos que estão sendo utilizados pelos usuários, já que a interação entre o bibliotecário e o usuário parece se dar cada vez mais através dos meios digitais de comunicação.

Um dos meios que pode ser utilizado para realizar este levantamento é analisar o rastro deixado pelos usuários após cada busca no sistema. Estes rastros são chamados de *logs* e também podem ser utilizados para verificar a estratégia de busca realizada pelos usuários.

Esta pesquisa utiliza as expressões de busca existentes nos *logs* coletados no portal do Senado Federal brasileiro chamado LexML. O portal LexML é um tipo de sistema de recuperação da informação especializado em Direito que fornece

acesso a diferentes tipos de documentos para os três poderes. Segundo Ferro (2014), o LexML possibilita o acesso a documentos e legislações para o Senado Federal, o Ministério da Justiça e a Câmara dos Deputados, entre outros órgãos federais, sendo a busca por termos indexados uma possibilidade que facilita o acesso a informação.

Os sistemas de recuperação da informação se tornaram parte do dia a dia do bibliotecário, por isso esta pesquisa busca resolver a seguinte questão: que tipos de variantes terminológicas do Direito Penal são utilizadas pelos usuários do Portal LexML no processo de recuperação de informação?

O objetivo geral desta pesquisa é analisar as tipologias de variação denominativa da área do Direito Penal utilizadas pelos usuários do LexML do Senado Federal Brasileiro para a recuperação da informação.

Este objetivo se divide em dois objetivos específicos que são:

- a) Identificar quais as tipologias de variantes denominativas na área do Direito Penal são utilizadas pelos usuários do Portal LexML;
- b) Verificar se as tipologias de variantes identificadas ocorrem no Vocabulário Controlado do Senado Federal Brasileiro.

A área do Direito deve ser acessível a todos, pois devemos compreendê-lo para convivermos melhor em sociedade. Sendo assim, se torna ainda mais importante se pensar as variações terminológicas para que se torne possível facilitar o acesso, tanto para o especialista, quanto para o leigo na área do Direito. Se o sistema de recuperação da informação não for consistente, documentos pertinentes podem acabar ficando fora da recuperação da informação.

Este trabalho se justifica pelas seguintes razões: a necessidade cada vez maior de se utilizar meios diferentes para diminuir as incompatibilidades nos sistemas de informação; a baixa quantidade de trabalhos realizados utilizando *logs* para a melhoria de linguagens controladas; o interesse da autora pelo tema desde as aulas de linguagem documentária III; e também a participação da autora no Projeto nº. 33016 – Análise da Variação Terminológica para Recuperação da Informação na área do Direito.

A Teoria Comunicativa da Terminologia de Cabré de (1998) é de grande importância para esse trabalho, assim como a tese de doutorado de Rita do Carmo Ferreira Laipelt (2015), o trabalho de conclusão de curso de Ferro (2014) e Faustich

(2006) e Freixa (2002) – isso quando falado nas variações. Lancaster (1979, 2004) é o principal autor utilizado quando o assunto é recuperação da informação, junto com Araújo (2012), Padilha e Café (2016), Passos e Caregnato (2018). Quando falamos em Indexação foram utilizado Lancaster (1979, 2004), Kuramoto (1995), Guedes e Dias (2010), Carneiro (1985), Olson e Boll (2001). Laipelt (2015) também foi utilizada para falar sobre a garantia do usuário, junto com Oliveira e Araújo (2012), Faustino (2014), entre outros.

## 2 REFERENCIAL TEÓRICO

Este capítulo apresenta o referencial teórico no qual o trabalho foi embasado, conceituando a Teoria Comunicativa da Terminologia, variação terminológica e os *logs* como forma de garantia do usuário.

### 2.1 Sistemas de Recuperação da Informação

A organização da informação é parte do dia a dia do bibliotecário. Através dela é que se possibilita que o usuário acesse à informação, ou seja, é através dela que se torna possível que o usuário recupere o documento.

Lancaster (1979) ressalta que o sistema de recuperação da informação é mais do que apenas a referência do documento, ele traz o resumo e a lista de termos que são os descritores do documento. Existem inclusive sistemas que recuperam o texto completo.

Para Araújo (2012), os sistemas de recuperação da informação possibilitam a representação, a organização e a localização dos itens de informação. Para que isso se torne possível, eles utilizam linguagens documentárias.

Segundo Laipelt (2015), estes sistemas de recuperação da informação são os intermediários entre o usuário e o acervo da instituição, sendo o bibliotecário então responsável por inserir os dados que representam o documento neste sistema. Ele é o responsável por pensar no usuário na hora de realizar o processo de indexação, que será abordado no tópico 2.2.

Sendo assim, os sistemas de recuperação da informação contam com um sistema de busca que possibilita que o usuário consulte o sistema. De acordo com Passos e Caregnato (2018), a solicitação feita pelo usuário é cruzada em um índice que pode ser de termos encontrados nos documentos, de autores, categorias e informação relacionadas.

Alguns destes sistemas podem ser customizados conforme a necessidade do local no qual ele está disponibilizado. Isto quer dizer que campos de descrição podem ser acrescentados ou retirados conforme se observe a necessidade ou não deles.

Como traz Fujita (2016), a organização da informação tem como objetivo uma rápida localização da informação. Percebe-se que essa organização da informação foi se modificando com o tempo devido ao avanço das tecnologias.

Neste trabalho o foco será nos descritores, já que serão estudadas as variações denominativas ocorridas durante as buscas no sistema e se os termos utilizados pelos usuários correspondem a linguagem utilizada durante a indexação.

Os descritores são os termos utilizados durante a busca, ou os termos preferidos para a indexação. É através deles que a recuperação da informação é possibilitada, seja ela por assunto ou por autor.

Os descritores fazem parte da linguagem de indexação, que segundo Fujita (2006), é a linguagem adotada para inserir dados de um documento dentro do sistema de recuperação da informação. Sendo esta sua relação com o sistema de recuperação da informação, pois este não funcionaria sem os descritores.

Fujita (2016) considera que na linguagem de indexação o processo de comunicação entre o sistema e o usuário deve ser facilitado. Ela diz que a linguagem utilizada no sistema deve ser a conhecida tanto pelo indexador quanto pelo usuário. Isto mostra a importância da garantia do usuário para um sistema consistente.

Estes sistemas podem ser manuais, automatizados ou mistos. Isto determinará se a extração e indexação dos termos será feita por um especialista, por um computador ou ambos.

Os registros nestes sistemas são feitos com a utilização de metadados. Se dois sistemas são compatíveis eles permitem a importação destes dados, o que segundo Padilha e Café (2016) torna possível que as buscas sejam ampliadas.

Os estudos terminológicos podem contribuir para uma melhor interoperabilidade entre sistemas. Porém cabe destacar que cada local é diferente, sendo assim, sempre será preciso adaptar a linguagem a seus usuários. Antes de importar dados de um sistema para o outro, é necessário observar cuidadosamente os descritores utilizados.

Os sistemas também contam com a possibilidade de execução de diferentes estratégias de busca. Segundo Passos e Caregnato (2018), o sistema de busca faz parte da arquitetura da informação. Com isso é possível perceber que a forma como o sistema está estruturado pode contribuir para a recuperação da informação.

Estes sistemas também podem e devem ser avaliados pelo bibliotecário. Durante estas avaliações, a indexação também seria avaliada, possibilitando assim que inconsistências no sistema, se houver, sejam corrigidas.

É possível perceber aqui a importância dos estudos terminológicos abrangendo a variação denominativa para a recuperação da informação, afinal o sistema de recuperação depende da indexação feita, e a indexação vai depender dos termos utilizados. Sem os estudos de variação estes termos podem ser diferentes dos utilizados pelos usuários e pode ocorrer de não haver remissivas.

## 2.2 Indexação

Muitos consideram que, com a tecnologia como parte essencial da vida das pessoas, o acesso à informação passou a ser facilitado, porém com o aumento do número de informações, ele de certa forma foi dificultado. Essa é apenas uma das mudanças que as tecnologias trouxeram para quem trabalha com informação. Segundo Guedes e Dias (2010), a preservação, guarda e acesso à informação sofreram diversas mudanças, afetando aqueles que trabalham com a informação. Porém parte das rotinas de representação da informação continuam iguais.

Representação da informação é um dos meios utilizados em bibliotecas para possibilitar uma posterior recuperação de determinados documentos, sendo estes documentos indexados em um sistema de recuperação da informação. Algumas das ações comuns na representação da informação são a leitura técnica e a extração do termo. A extração do termo pode ser realizada por um especialista ou por um computador, no caso dos sistemas automatizados. Durante este processo é utilizada a leitura técnica. A leitura técnica consiste na leitura das partes que podem fornecer maiores informações sobre o documento.

A representação da informação está relacionada com a indexação devido ao fato de a indexação ser parte da representação da informação. Sendo a indexação apenas um dos processos que são realizados nas rotinas de representação da informação.

De acordo com Lancaster (2004), devido ao fato de o bibliotecário nem sempre poder ler o documento que irá indexar do início ao fim, as partes que devem ser lidas mais cuidadosamente são título, resumo, sinopse e conclusões. Já a NBR

12673 (1992) acrescenta a estes a introdução, ilustrações, diagramas, tabelas e seus títulos explicativos, palavras ou grupos de palavra em destaque e as referências bibliográficas.

Conforme a NBR 12673 (1992) para atribuir os termos, todo o documento deve ser considerado pelo bibliotecário. Sem a atribuição dos termos e demais informações o documento não poderia ser recuperado posteriormente. Demais informações seriam o autor, ano de publicação/divulgação e local quando houver, que também são essenciais na descrição do documento dentro do sistema.

Lancaster (1979) também traz o processo feito antes de o documento ser incorporado ao acervo. Esses processos são a classificação, catalogação, indexação e resumos. Estas atividades, se executadas corretamente, possibilitam uma posterior recuperação da informação mais eficaz e mais rápida.

Quando faz a busca, o usuário não sabe qual termo foi utilizado e pode acabar não recuperando nada caso não sejam utilizadas remissivas. Isso deve-se ao grande número de variantes que podem ser utilizadas e da possibilidade de se utilizar um termo genérico ou um específico no processo de busca. Há também a possibilidade de termos diferentes terem sido utilizados para se referir a um mesmo assunto tornando o sistema inconsistente. Por isso as palavras são importantes na hora da recuperação da informação.

Segundo Kuramoto (1995), as causas para que um sistema não retorne um grande número de informações relevantes são diversas, entre elas estão as estratégias de busca, a terminologia utilizada, já que uma palavra pode ter sentidos diferentes conforme a área do conhecimento a qual pertence, as mesmas palavras podem ser usadas com sentidos diferentes e palavras completamente diferentes podem ser utilizadas para exprimir o mesmo conceito.

Segundo Guedes e Dias (2010), a representação da informação ocorre em dois níveis: o primário e o secundário. No primário a informação produzida pelo é produto dos pensamentos dele e no secundário é o que vai para os sistemas de informação, para que o documento seja encontrado facilmente.

Quando vai ser acrescentado ao sistema, o documento precisa de uma análise e depois uma tradução. Na análise é feita a leitura técnica e é onde se define o assunto do documento. A tradução consiste na escolha dos termos que irão representar aquele documento no sistema.

Segundo Lancaster (1979), o primeiro passo é a análise que determina o que o usuário estaria possivelmente procurando quando busca aquele documento e o segundo passo é a tradução quando essa constatação se transformará em um vocabulário. Neste contexto é feita uma distinção entre indexação por extração, que Lancaster (2004) afirma ser aquela na qual os termos utilizados estão no texto, e indexação por atribuição, onde os termos estão em uma fonte que não é o documento.

Esses dois processos fazem parte da indexação, que nada mais é do que os procedimentos executados para representar a informação. A indexação pode ser feita por um especialista, geralmente o bibliotecário, ou de forma automática que é feita por um sistema especialmente programado para isso. Ambas têm suas vantagens e desvantagens, sendo assim o melhor seria utilizar ambas para acrescentar o olhar humano à indexação. Entre as vantagens da indexação feita por um especialista, está a profundidade maior na indexação. O especialista consegue pensar nas possíveis necessidades que aquele documento pode responder, o que na automática não é possível.

Por isso Vale (1987) a escolha da linguagem de indexação é de extrema importância. É através dela que o sistema se tornará ou não eficaz. Se não houver um cuidado na escolha da linguagem de indexação, diferentes termos podem ser utilizados para se referir ao mesmo assunto, ou remissivas podem acabar ficando de fora. Segundo Wallauer (2009), remissivas são termos não autorizados que remetem a termos autorizados. Segundo Santiago e Krieger (2008), estes termos compõem uma rede, chamada de rede de remissivas, que possibilita a recuperação da informação. Para Fujita (2016), as linguagens de indexação são linguagens artificiais utilizadas para representação da informação durante seu tratamento temático.

De acordo com Fujita (2016) existem dois tipos de linguagens de indexação a alfabética e a hierárquica – sendo os cabeçalhos de assunto e os tesouros exemplos de linguagens alfabéticas e a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU) exemplos de linguagem hierárquica. Como o foco do trabalho são os termos, as linguagens hierárquicas não serão aprofundadas.

Nem sempre uma linguagem controlada (cabeçalho de assuntos e tesouro) é utilizada para indexação.



Segundo Fujita (2016), as linguagens utilizadas na indexação alfabética podem ser pré ou pós coordenadas. No caso da pré coordenada a coordenação é feita na hora da indexação, já na pós coordenada isto ocorre na estratégia de busca.

A indexação nos sistemas de recuperação da informação pode ser feita através de linguagem natural, que é a linguagem extraída diretamente do texto, ou de linguagem controlada, que é quando o termo é retirado de um tesouro ou cabeçalho de assuntos, por exemplo. Também há a linguagem livre, onde qualquer termo considerado relevante, estando no texto ou não, pode ser acrescentado.

Segundo Lancaster (2004), há uma tendência maior à utilização de vocabulários controlados. Um vocabulário controlado seria uma lista de termos autorizados pela instituição. Estes vocabulários, para Lancaster (2004), são classificações bibliográficas, cabeçalhos de assuntos e tesouros, sendo que os cabeçalhos de assunto apresentam problemas maiores por não distinguir claramente as relações hierárquicas das associativas. Porém, comumente esses vocabulários são também uma forma de estrutura semântica controlando sinônimos, diferenciando homógrafos e reunindo termos relacionados. Segundo Fujita (2016) entre os vocabulários controlados, os tesouros são mais utilizados. Talvez por causa dos problema que existem no cabeçalho de assunto.

Porém, na Terminologia é entendido que raramente se tem sinônimos dos termos, o que são apresentados por essas linguagens na maioria das vezes são as variações. Também cabe destacar que nem todas as áreas do conhecimento possuem um vocabulário alfabético controlado.

Infelizmente, os termos utilizados para indexação podem não ser os mesmos que são utilizados nas buscas. Isto ocorre principalmente em bibliotecas onde o público é mais variado, pois elas atendem tanto a leigos quanto a especialistas. Por isso é necessária uma avaliação cuidadosa do vocabulário controlado utilizado. Estes vocabulários também precisam de uma avaliação para que sejam atualizados sempre que necessário.

O indexador deve ter em mente o usuário na hora de fazer a tradução do documento, pois como será visto, as diversas áreas do conhecimento têm diversas variações de seus termos ou podem utilizar o mesmo termo com significado completamente diferentes.

Para Lancaster (2004), ao realizar uma indexação trazendo assim aquele documento para o acervo, se espera que quando a necessidade surgir, o usuário possa recuperar o máximo de documentos pertinentes.

Se não pensar no usuário, quando fizer a indexação, o bibliotecário pode tornar o sistema inconsistente, recuperando assim poucos materiais relevantes durante a busca. Alguns termos possuem uma tendência maior a serem procurados do que suas variantes, sendo esses os que devem ser utilizados como descritores.

Segundo Lancaster (2004), a cobertura da base de dados, sua política de indexação, sua política e prática na redação de resumos, a qualidade do vocabulário utilizado na indexação entre outras coisas determinarão o sucesso ou não na recuperação da informação.

Os usuários podem recorrer a diversos tipos de bibliotecas com suas necessidades de informação, que nem sempre são claras para eles. Segundo Lancaster (1979) essas necessidades podem ser: localizar e obter uma cópia de um documento do qual já conhecem o título e o autor, ou localizar documentos com um assunto em particular. O tipo de biblioteca determinará o usuário, que por sua vez ajudará a definir que termos utilizar. Por exemplo, um artigo sobre direito à saúde pode vir a ser indexado em uma biblioteca específica do Direito, assim como pode vir a ser indexado em uma biblioteca médica, porém ambas as indexações utilizarão termos diferentes, pois as necessidades de informação são diferentes.

É como exemplifica Lancaster (2004): um artigo da Nasa na biblioteca da Nasa pode vir a ter descritos todos os aspectos interessantes para essa biblioteca, porém o mesmo artigo também pode ser de interesse dos usuários de uma fábrica de borracha devido a confecção dos trajes dos astronautas serem desse material. Entretanto, para os usuários da fábrica de borracha será necessário usar termos mais específicos sobre a borracha. Por isso, o autor traz que a indexação deve ser ajustada aos interesses do grupo atendido pela unidade de informação.

Os interesses do grupo também podem se modificar com o tempo, modificando assim o modo de uso da coleção. Segundo Lancaster (2004), isso fica visível através dos relatórios de pesquisa dentro de instituições. Por isso o indexador deve também estar atento a possíveis usos futuros. Neste sentido, os *logs* também podem ser considerados um tipo de relatório de pesquisa.

O tipo de indexação vai variar conforme a biblioteca, assim como a linguagem utilizada. Por isso é muito importante que também haja uma política de

indexação na biblioteca. Nela estarão registradas as decisões tomadas quanto ao tipo de linguagem, número de descritores, entre outras.

De acordo com Fujita (2016) a política de indexação são as decisões tomadas que delimitam a cobertura temática e demais procedimentos relativos ao sistema de recuperação da informação. Algumas destas decisões são relativas a indexação tais como: a linguagem de indexação, a exaustividade entre outras.

Para Carneiro (1985), os elementos que devem fazer parte de uma política de indexação são: cobertura de assuntos, seleção e aquisição, processo de indexação, nível de exaustividade, nível de especificidade, escolha da linguagem, capacidade de revocação do sistema, estratégia de busca e avaliação do sistema.

Para Lancaster (2004) uma das decisões mais importantes quando se elabora uma política de indexação é a exaustividade, que é uma estimativa de quantos termos serão utilizados na indexação do documento, porém não deve ser um limite absoluto.

Não desconsiderando a importância da exaustividade, para a autora do presente trabalho a decisão mais importante é a linguagem utilizada. Para Olson e Boll (2001) a adequação, exaustividade, especificidade e consistência são elementos que afetarão a busca e a recuperação da informação.

A indexação é complexa, pois é afetada pela terminologia e pelos usuários, assim como pela biblioteca na qual o conjunto de documentos está inserido. A indexação também depende do país ou cidade em que o indexador está, devido a normas culturais e linguagens divergentes. A indexação afeta os demais serviços da biblioteca, como o serviço de referência entre outros. Sendo assim, a terminologia afeta todos os serviços desenvolvidos na biblioteca.

O indexador deve ter consciência da importância do trabalho que está realizando e de que, a forma como ele realiza, este trabalho influenciará na satisfação do usuário. Essa satisfação, conseqüentemente, terá impacto em como a biblioteca é vista, podendo trazer mais usuários para ela ou afastar os que já utilizaram ou utilizam o serviço.

## 2.3 Variação denominativa

A variação faz parte do discurso especializado escrito e falado. Ela possibilita a quem escreve utilizar diferentes termos para expressar o mesmo sentido. Claro que alguns termos algumas vezes são mais adequados do que outros.

A variação é estudada na Terminologia, que segundo Cabré (2008), é interdisciplinar, pois está presente nas diversas especialidades e tem suas bases em outros campos do saber. Ela busca fornecer uma comunicação clara e sem a possibilidade de mal entendidos entre os profissionais.

Cabré (2009) diz que a partir da segunda metade do século XX, as trocas sociais, econômicas e tecnológicas facilitaram o reconhecimento da Teoria da Terminologia (TGT). Ou seja, viu-se a importância que a Terminologia tem para a sociedade. De fato foi se notando maior necessidade da Terminologia.

Cabré (2009) relata que, com a observação dos termos do discurso, algumas críticas a Teoria Geral da Terminologia começaram a surgir, pois perceberam que ela considerava os aspectos fonéticos e sintáticos do léxico. A TGT também não descreve bem a separação de termos e palavras. Para Cabré (2009), uma teoria adequada deve supor a explicitação de seu objeto de estudo, a representação deste objeto em dados e respeitar a interdisciplinaridade que existe na Terminologia.

Dentro da Teoria Comunicativa da Terminologia, Cabré (2009) descreve os termos como unidades comunicativas que são unidades de discurso, que fazem parte normalmente de áreas especializadas, apresentam variações (denominativas e conceituais). A variação denominativa frequentemente tem consequências cognitivas e a variação conceitual ocorre por diferentes critérios, entre eles a escola ideológica, o ponto de vista, entre outros. Para Cabré (2009) os termos compartilham unidades de outros sistemas da denominação de conceitos especializados e compartilham com as unidades não linguísticas a transmissão de informação especializada.

Cabré (2009) relata que desde 1996 trabalha com a Teoria Comunicativa da Terminologia, a qual descreve como uma teoria linguística das unidades terminológicas e de base cognitiva com propósito comunicativo. Esta teoria pressupõe as unidades terminológicas como objeto central da Terminologia, como campo de conhecimento, e essas unidades são linguísticas, cognitivas e sociocomunicativas.

A autora (2009) também aborda em sua teoria que existem diversas portas para o acesso às unidades terminológicas e que cada porta exige uma teoria própria

que deve ter como objetivo central as unidades terminológicas e sua concepção poliédrica.

As variações ocorrem nos termos que o bibliotecário pode utilizar como descritores. Para Faulstich (2006) os termos são signos que se tornam funcionais na linguagem. Também fazem parte da linguagem especializada.

Segundo Faulstich (2006), o próprio termo é variante, pois dentro do discurso especializado ele vai assumir uma função de acordo com o contexto do uso. É válido aqui acrescentar que este contexto pode também se aplicar a região, já que em alguns casos em locais do mesmo país termos tem um significado completamente diferente.

Segundo Cabré (2008) termo e conceito estão interligados, sendo que o termo é a unidade utilizada para designar um conceito, que seria a representação consensual de um objeto da realidade, pois a percepção do conceito vai variar de acordo com o grupo cultural.

Por isso um conceito esta interligado a área do saber a qual pertence, sendo assim o termo também. Para Cabré (2008), um conceito está sempre associado ao âmbito temático da representação. Neste trabalho a esfera temática é o Direito Penal. De acordo com Cabré (2008), necessidades distintas criam percepções distintas, assim como o conhecimento adquirido ao longo do tempo. Assim pressupõe-se que o especialista tenha adquirido o conhecimento necessário para compreender o discurso da área.

A partir da tese *“Pour une socioterminologie – des problèmes sémantiques aux pratiques institutionnelles”* de Gaudin (1993), a variação terminológica passou a ser pesquisada e sua importância veio à luz. Segundo Laipelt (2015) os primeiros passos no estudo da variação se deram a partir desse trabalho de Gaudin que falava da análise da circulação dos termos, ou seja de Socioterminologia. Sendo assim é possível perceber que os estudos de variação terminológica são anteriores a Teoria Comunicativa da Terminologia.

A variação terminológica é uma das principais causas do pouco retorno ou do retorno excessivo das respostas de buscas. Por esse motivo ela é muito importante na Biblioteconomia e influenciam diretamente na indexação e consequentemente, na recuperação da informação.

Devido ao grande número de variações é necessário um cuidado maior na hora da indexação. Como, em muitos casos, bibliotecários diferentes trabalham

indexando obras no mesmo local, se não houver um vocabulário controlado eles podem acabar inserindo termos diferentes para um mesmo assunto. Essas diferenças acabam tornando o sistema inconsistente e gerando um baixo retorno na hora em que a busca é efetuada.

Segundo Cabré (2008), o especialista tem conhecimento dessas variações, mesmo que ele não perceba. Porém cabe aqui ressaltar que, mesmo tendo esse conhecimento, o especialista não vai ter certeza de qual o termo utilizar ao realizar a busca. Isso pode levar a necessidade de diversas pesquisas, o que pode ser evitado com o uso de uma linguagem controlada adequada. Cabré (2008) resalta que a interpretação do texto pode ser influenciada pela variante utilizada. Por isso, considerar a variação se torna crucial dentro da Biblioteconomia – mesmo que a utilização de linguagens controladas seja pequena e que existam poucos tesouros especializados disponíveis.

É possível que um usuário especialista e um usuário leigo utilizem diferentes termos para se referir a mesma coisa, como competência da justiça do trabalho e responsabilidade da justiça do trabalho. Ambos certamente com diferentes objetivos e também eles com um retorno de diferentes documentos, se não houver uma remissiva no catálogo.

Isso também ocorre na área do Direito Penal, onde uma pessoa que não é especialista na área na área pode utilizar determinado termo e um especialista utilizar outro completamente diferente. Como o Direito é uma área que deve ser acessível a todos, isso se torna ainda mais relevante.

Geralmente são considerados dois tipos de variação terminológica: a denominativa e a conceitual. Segundo Laipelt (2015) a variação denominativa se refere aos diferentes termos que representam um mesmo conceito e a variação conceitual se refere a um termo com diferentes conceitos.

Este trabalho será focado apenas na variação denominativa devido ao tempo disponível para o desenvolvimento da pesquisa. A variação denominativa, segundo Freixa (2014, *apud* Laipelt, 2015) pode ocorrer através de alterações gráficas, alterações morfosintáticas, reduções, alterações léxicas de unidades poliléxicas e alterações complexas. Tudo isso influencia na hora da recuperação da informação.

De acordo com Freixa (2002) as alterações gráficas se dividem em termo e sua forma artificial, termo e suas abreviações, alterações ortográficas. O tipo de alteração gráfica classificado como termo e sua forma artificial se subdivide como:

termo e símbolo, termo e sua fórmula química e também como termo e outra forma artificial.

Freixa (2002) apresenta dois casos em que a variação seria classificada como termo e abreviação, são eles o termo e sigla, e o termo e sua forma abreviada. Para as alterações ortográficas, Freixa (2002) divide como presença e ausência de acentuação, presença e ausência de hífen, erros de grafia, presença ou ausência de espaço entre as palavras. Exemplos dos subtipos de alterações gráficas serão apresentados no quadro 1 abaixo.

**Quadro 1 - Alterações gráficas**

<b>Tipo</b>	<b>Subtipo</b>	<b>Exemplos</b>
Termo e forma artificial	Termo e símbolo	Cobre/CU
	termo e fórmula química	Aminíaco/ $\text{NH}_3$
	termo e outra forma artificial	Nova lei dos motoristas/Lei 12.619/2012.
Termo e abreviação	Termo e sigla	Carteira de trabalho e previdência social/CTPS
	Termo e abreviação	Horas Extras/Hrs extras;
Alteração ortográfica	- presença e ausência de acentuação	Férias/ férias
	- presença e ausência de hífen	Auxílio-alimentação/ auxílio alimentação
	- erros de grafia como acréscimo ou supressão de letras	Paradigma diferença salarial/paradiguima diferença salarial
	- presença ou ausência de espaço entre as palavras	Contracheque/ contra cheque

Fonte:Laipelt (2015)

De acordo com Freixa (2002), as variações morfossintáticas podem se aplicar ao substantivo ou a preposição, ou seja, elas podem ser determinadas pela

existência ou não da preposição. São divididas em duas: manutenção da estrutura e alteração da estrutura. No primeiro caso não há uma modificação na estrutura.

Elas estão na estrutura morfossintática do termo que se mantém igual no caso de ocorrer no artigo, no nome ou no gênero, e se modifica ela ocorrer da forma [Substantivo + adjetivo] / [Substantivo + Sintagma preposicional] onde um deles utiliza a forma considerada culta e o outro a forma popular.

Algumas variações podem se apresentar de forma monoléxica ou poliléxica, que corresponde ao uso de uma ou mais palavras para um mesmo conceito, como o caso de produto ecológico cuja a variante pode ser ecoproduto. Neste caso, Laipelt (2015) diz que a alteração entre estruturas monoléxicas/poliléxicas o prefixo é representante do adjetivo.

Uma das variações que ocorre no “Direito do Trabalho” é a do termo “licença maternidade” o qual algumas vezes é utilizado como licença gestante. Essa variação seria o que Freixa (2002) classifica como outros tipos de alterações da estrutura.

Laipelt (2015), traz como exemplo de alteração morfossintática do subtipo alteração do substantivo “execução trabalhista contra herdeiro/execução trabalhista contra herdeiros”. Neste caso fica claro a alteração no substantivo herdeiro que passa para o plural.

## Quadro 2 – Alterações morfossintáticas

Tipo	Subtipo	Exemplos
Manutenção da estrutura	Ausência/ presença de artigo	Acidente de trabalho/acidente do trabalho; visto que “do” é resultado da contração da preposição “de” + artigo definido “o”
	Alteração do substantivo	Execução trabalhista contra herdeiro/ execução trabalhista contra herdeiros
	Alteração de preposição	Acidente em serviço/acidente de serviço;
	Alteração do gênero	Empregada doméstica/empregado doméstico.



Alteração da estrutura	[substantivo + Adjetivo]/ [substantivo + sintagma preposicional]	Vínculo empregatício/ vínculo de emprego; neste caso o adjetivo geralmente é a parte alterada do termo
	Monoléxico/ poliléxico	Produto ecológico/ecoproduto
	Outras alterações de estrutura	Primeira matéria/ matéria primeira

Fonte: Laipelt (2015)

O próximo tipo de variação denominativa são as reduções. Laipelt (2015) diz que há duas categorias dentro das reduções: as reduções de extensão e as reduções de base. Estas categorias são divididas em seis, três em cada. A primeira categoria se divide em [Substantivo + X] podendo X ser um adjetivo. . [substantivo + x + y] podendo x e y se referirem a um adjetivo ou a um sintagma preposicional. É possível observar as reduções de extensão no quadro 3 e as reduções de base no quadro 4 abaixo.

### Quadro 3 – Reduções de extensão

Subtipo	Exemplos
a) Tipo [substantivo + x] = [substantivos]	
- [substantivo + adjetivo]	Aviso prévio/ aviso;
- [substantivo + sintagma preposicional]	Acidente de trabalho/acidente;
- [substantivo 1 + substantivo 2]	Gratificação assiduidade/gratificação
b) [substantivo + x + y] = [substantivo + x]	
- [substantivo + sintagma preposicional + adjetivo] = [substantivo + sintagma preposicional]	Equipamento de proteção individual/equipamento de proteção;
- [ substantivo + adjetivo + sintagma preposicional] = [substantivo + adjetivo]	Certidão negativa de débitos/certidão negativa

- [ substantivo + adjetivo 1 + adjetivo 2] = [substantivo + adjetivo 1]	Certidão negativa trabalhista/certidão negativa
- [substantivo + sintagma preposicional 1 + sintagma preposicional 2] = [substantivo + sintagma preposicional 1]	Adicional periculosidade de eletricista/adicional periculosidade
C) Outras reduções de extensão	Tabela de lotação de cargos/tabela

Fonte: Laipelt (2015)

#### Quadro 4 – Reduções de Base

Subtipo	Exemplos
d) [Substantivo + Adjetivo] = [Adjetivo N]	Trabalhador temporário/ Temporário
a) [ Substantivo 1 + substantivo 2] = [substantivo 2]	Adicional Insalubridade/ Insalubridade
b) [substantivo 1+ sintagma preposicional (de + substantivo 2)] = substantivo 2	Pagamento de diárias/ diárias
c) Outras reduções	Trabalho Doméstico Infantil/ Trabalho Infantil

Fonte: Laipelt (2015)

De acordo com Laipelt (2015) as alterações léxicas de unidades poliléxicas se dividem em dois tipos: alteração de base e alteração de extensão. Sendo que as alterações de base se dividem em três subtipos e as alterações de extensão também. Freixa (2002) divide as formas poliléxicas de acordo com a parte do sintagma (base ou extensão) em que está a mudança e a estrutura morfossintática do sintagma. Como mostrado no quadro 5 a seguir.

#### Quadro 5 - Alterações léxicas de unidades poliléxicas

Subtipo	Exemplos
a) Alteração de base	
- [substantivo+ sintagma preposicional (pre.+ art.) + substantivo]	Jornada de trabalho/ horário de trabalho
- [substantivo + adjetivo]	Empregado doméstico/ trabalhador doméstico
- [Outras alterações de base]	Competência da justiça do trabalho/ responsabilidade da justiça do trabalho
b) Alterações de extensão	
- [substantivo+ sintagma preposicional (pre.+ art.) + substantivo]	Acidente de trabalho/ acidente de serviço;
- [substantivo + adjetivo]	Abono salarial/ abono pecuniário
- [Outras alterações de extensão]	Consolidação das leis do trabalho/ consolidação das leis trabalhistas

Fonte: Laipelt (2015)

As várias alterações complexas, segundo Laipelt (2015), estão organizadas de acordo com o parentesco formal ou não entre as variantes analisadas. Sendo

assim elas se subdividem em com parentesco formal e sem parentesco formal e essa se subdividem em monoléxica/ poliléxica, como no quadro 6 abaixo.

Quadro 6 – Várias alterações complexas

Tipo	Subtipo	Exemplo
Com parentesco formal	a) Monoléxica/ poliléxica	Agroecologia/ Agricultura ecológica
	b) Poliléxica/ poliléxica	Doenças ocupacionais/ Doenças profissional
Sem parentesco formal	c) Monoléxica/ poliléxica	Sobreaviso/ tempo à disposição
	d) Poliléxica/ poliléxica	Bombeiro Civil/ Brigadista particular

Fonte: Laipelt (2015)

Como é possível observar, as variações denominativas são ocorrências comuns e isso faz com que seja ainda mais importante estudá-las. Na área da Biblioteconomia elas são de extrema importância na recuperação da informação. Cabré (2009) nota um aumento na necessidade de estudar assuntos ligados à língua, principalmente no âmbito especializado. Lembrando que as variantes são analisadas no mínimo entre pares, pois é necessário verificar um termo em relação ao outro(s).

## 2.4 Análise de Logs

Os *logs* são rastros deixadas pelo usuário no momento em que faz a pesquisa, que podem ser visualizados. Eles fornecem uma visão do caminho percorrido pelo usuário, que pode ser utilizada para fazer melhorias no sistema. Porém no Brasil muitas instituições ainda não fazem a guarda de *logs*, pois dizem que ocuparia muito espaço no servidor.

Segundo Nicholas, Huntinggton e Watgkinson (2005), os *logs* podem prover uma informação mais real sobre a busca dos usuários, pois não vêm de uma análise anterior e sim diretamente da busca feita pelo usuário. Todo o sistema de recuperação da informação pode facilmente contar com os *logs*, desde que esteja configurado para a guarda deles.

Os *logs* trazem como informação muitos dados, pois na verdade é o registro de toda e qualquer atividade realizada pelo usuário no sistema. Alguns destes dados podem ser úteis para pesquisas de variação terminológica, como o IP do computador, data, hora, minutos, segundos, o país de onde a pesquisa foi realizada,

a expressão de busca e se teve sucesso ou não na recuperação da informação. São as expressões de busca que serão analisadas no presente trabalho.

Os *logs* não deixam de ser uma fonte vital de informação para aqueles que estudam a terminologia. No caso da Biblioteconomia ele virá a contribuir para a elaboração de linguagens documentárias alfabéticas, tais como cabeçalhos de assunto e tesouros.

Eles também podem vir de diversas áreas, apesar do trabalho se concentrar nos *logs* do Direito. E apesar da maioria das bases de dados fornecerem busca a busca por autor, há uma tendência a se realizar buscas por assunto.

Os *logs* também podem ser utilizados no estudo de usuários, para verificar quais assuntos são mais procurados, assim contribuindo na avaliação do acervo, porém eles ainda são mais utilizados para análises terminológicas. Essas análises ajudam a reduzir os resultados não relevantes nas buscas, o que em grande parte das bases não acontece. Sendo assim, se tem um retorno excessivo e a necessidade de um refinamento nas buscas.

É muito comum o usuário receber uma grande quantidade de documentos quando realiza buscas em Sistemas de Recuperação de Informação (SRI), sendo necessário refinar os termos usados na consulta objetivando reduzir o número de documentos para leitura e análise. Porém o refinamento da estratégia de busca faz com que muitos documentos importantes acabem sendo omitidos, ficando invisíveis para o usuário. Os tradicionais SRI são normalmente baseados em um modelo reducionista, utilizando a lógica booleana para identificação de documentos[...] (GABRIEL JUNIOR, 2016, p.1)

Essa grande quantidade de informação recebida pelo usuário é um dos fatores que influenciam na utilização dos *logs*, pois através deles é possível observar se há inconsistência entre o que se pesquisa e o termo indexado.

No caso de inconsistências no sistema através dos *logs*, referentes às expressões de busca realizadas no sistema de recuperação da informação, é possível ver qual o termo mais procurado. Possibilitando assim colocar aquele como o descritor e os outros como remissivas. Eles também podem ser uma fonte consultada na hora de elaborar um vocabulário controlado.

Os vocabulários controlados ajudam a manter a consistência no sistema. Se durante a elaboração de um vocabulário controlado, os *logs* forem consultados, a linguagem ficará o mais próxima possível da utilizada pelo usuário.

A análise dos *logs* de busca pode contribuir inclusive para a inclusão de novas remissivas, já que a linguagem está sempre se atualizando. Novas palavras surgem, algumas entram em desuso. Porém apesar disto guardar os *logs* ainda não é uma prática comum nas instituições do Brasil.

Sendo assim ainda não se tem muitos estudos que utilizem eles como fonte de informação. Muitos lugares ainda consideram que os *logs* ocupam muito espaço, parecendo esquecer as grandes contribuições que eles podem trazer para áreas como a da terminologia e da biblioteconomia.

Utilizando os *logs* é possível realizar inclusive um estudo dos sistemas, contribuindo assim para a área da informática. Isto se deve ao fato de que através dos *logs* é possível ver toda a jornada do usuário durante as suas buscas.

De acordo com Job, Mattos e Ferreira (2013), o registro nos logs de busca pode ajudar a quantificar o acesso a artigos de periódicos. Sendo assim os editores podem analisar e disponibilizar dados como número de acesso e local onde determinado artigo é acessado com maior frequência.

Segundo Job, Mattos e Ferreira (2013), se tabulados, os *logs* podem fornecer informações como autoria, tipologia e mais acessados. Isto pode contribuir para o acréscimo ou não de publicações de determinado autor no acervo da biblioteca. Pois podem fornecer ao bibliotecário quais autores estão sendo mais lidos em cada área.

## 2.4 Garantia do usuário

A garantia do usuário faz parte de um processo que influencia diretamente na indexação, pois ela é parte fundamental da elaboração de linguagens controladas. Sendo que de acordo com Moreira e Moura (2006) ela é o procedimento através do qual o bibliotecário torna o usuário uma fonte que contribuirá para indexação.

Este processo ajuda a tornar a linguagem mais próxima a utilizada pelo usuário já que, cada área do conhecimento possui um vocabulário específico. Também é entendido que considerar apenas os textos das áreas do conhecimento

não garantiria que a linguagem utilizada para indexação seria a mesma utilizada pelo usuário depois. Como trazem Moreira e Moura (2006) os usuários são distintos com necessidades diferentes, o que demonstra a importância desses processos durante a elaboração de uma linguagem controlada ou até mesmo para analisar a consistência do sistema de recuperação da informação.

Como todas as áreas, o Direito tem termos específicos, como *habeas corpus*, o que, para algumas pessoas, dificulta o acesso à informação necessária, pois não são termos geralmente escutados no dia a dia.

A elaboração de tesouros é parte do trabalho do bibliotecário. A elaboração dessas linguagens possibilitam que a indexação se torne mais precisa, ajudando assim numa recuperação satisfatória da informação. Na elaboração desse tipo de linguagem é constituído um *corpus textual* que servirá como garantia literária.

Para a construção desse *corpus*, são estudados e coletados os textos da área específica de conhecimento para a identificação e extração de termos e conceitos do domínio. No intuito de garantir que a linguagem utilizada pelos usuários ao expressarem suas necessidades de busca coincidam com a linguagem utilizada pelos bibliotecários no processo de indexação, utilizam-se os *logs* das buscas ou as pesquisas que estão sendo realizadas na área abordada pelo tesouro.

Mesmo que o *corpus* seja especialmente selecionado, o que deve ser, ainda assim não se tem a garantia do usuário, apenas a garantia textual. A garantia do usuário é essencial como destacam Oliveira e Araújo (2012)

No processo de construção de um tesouro, alguns aspectos devem ser considerados a fim de obterem-se resultados que possam legitimá-lo perante o campo específico para o qual esteja sendo desenvolvido. Dentre eles, destaca-se, além da garantia literária e da garantia estrutural, a garantia do usuário, tendo em vista que este deverá ser o maior beneficiário desta ferramenta de recuperação da informação. (OLIVEIRA; ARAÚJO, 2012, p.23)

Quanto mais automatizado os sistemas se tornam, mais o bibliotecário precisa encontrar formas para a garantia do usuário durante a indexação. Como vimos acima, as variações podem ser um problema na hora do resultado da busca, caso o bibliotecário não leve em conta o usuário, quando seleciona o termo para a indexação ou até mesmo para a elaboração de um tesouro.

Com a diminuição do acesso presencial dos usuários nas bibliotecas, fica cada vez mais complicado obter a garantia do usuário, que é como é chamada a pesquisa sobre a forma como o usuário faria a busca. Este meio é utilizado quando se vai elaborar uma linguagem documentária.

Segundo Gabriel Junior e Laipelt (2017), os tesouros são um auxílio na hora da indexação, porém na sua elaboração as relações utilizadas ainda são as simples, baixando assim a relação semântica, pois não se especifica o tipo de relação entre o descritor e os termos relacionados. Porém, seu uso ainda é de extrema importância para garantir a precisão, embora eles ainda não sejam linkados com o sistema de recuperação da informação, ou seja, nos atuais sistemas de recuperação a informação, na maioria dos casos ainda não é possível acessar o tesouro direto do sistema, porém ele deve ser disponibilizado também para os usuários.

Por mais que tenham diferentes interfaces, os sistemas de recuperação da informação têm em comum o campo de busca, é lá que são inseridos os termos pesquisados.

Assim, como é possível saber os termos de busca utilizados sem ter um contato direto com o usuário? É neste aspecto que os *logs* vêm contribuir com os estudos terminológicos, e mais especificamente, para a construção de recursos de organização do conhecimento, visto que, de acordo com Fujita (2008), a organização do conhecimento é o conjunto de processos feitos para representar e possibilitar o acesso aos documentos. Os logs, portanto, podem servir como fonte de informação na hora de elaborar um tesouro, o que ajudará na indexação.

Ao analisar a importância da temática da garantia do usuário, encontram-se, na literatura, termos similares, como “garantia de uso”, ou “endosso do usuário”. Na construção das linguagens documentárias, observa-se que esses termos vinculam-se fortemente às linguagens de indexação que se apresentam como instrumentos indispensáveis à recuperação da informação. (OLIVEIRA; ARAUJO, 2012, p.23)

Com uma análise de *logs*, o bibliotecário pode observar quais termos estão sendo utilizados e definir melhor qual o termo preferido, ou seja, qual o termo será o principal e quais serão os termos equivalentes. Os termos não preferidos podem e devem se tornar remissivas, contribuindo para a satisfação do usuário com uma

resposta melhor do sistema. Claro que nem todos os sistemas de recuperação da informação possibilitam o uso de remissivas, mas, sempre que houver esta opção as remissivas devem ser utilizadas.

Segundo Oliveira e Araújo (2012), embora seja muito importante, ainda são poucos os casos que retratam como os bibliotecários as têm utilizado para elaboração de Tesouros e outras linguagens controladas.

Utilizando a garantia do usuário para fazer uma análise da indexação no sistema pode melhorar a resposta. Também pode se fazer melhorias na linguagem controlada através desta análise.

Através destas informações é possível ver a importância da garantia do usuário e também de começar a se guardar os *logs*, mesmo que seja por um período de tempo pré-estabelecido em alguma política da biblioteca.



### 3 METODOLOGIA

Está é uma pesquisa básica já que não foi feita com o intuito de ser aplicada diretamente, mas sim com o intuito de conhecer como o usuário expressa suas necessidades de informação a partir da análise dos termos por ele utilizados. Segundo Gerhard e Silveira (2009) este tipo de pesquisa contribui para o avanço da ciência, visto que ao ser realizada, traz novas descobertas e agrega conhecimento ao que já se sabe.

A abordagem desta pesquisa foi qualitativa, pois trabalhou com as linguagens de busca e indexação para análise em profundidade dos casos de variação denominativa mais frequentes identificados no *corpus* da pesquisa.

Quanto ao procedimento se caracteriza como pesquisa documental, pois os *logs*, apesar de não serem o que no senso comum se entende por documento são assim considerados para pesquisas. Os log também não receberam um tratamento analítico anterior. De acordo com Godoy (1995), novos conhecimentos podem ser inseridos após uma análise dos documentos. Os *logs* não são por ela mencionados por ser uma publicação de 1995, mas suas características ficam evidentes na caracterização de pesquisa documental feita pela autora.

Como comumente pensamos que o trabalho de pesquisa sempre envolve o contato direto do pesquisador com o grupo de pessoas que será estudado, esquecemos que os documentos constituem rica fonte de dados. O exame de materiais de natureza diversa, que ainda não receberam um tratamento analítico, ou que podem ser reexaminados, buscando-se novas e/ ou interpretações complementares, constitui o que estamos denominando pesquisa documental. (GODOY, 1995, p.21)

Como é possível observar através da citação, o tipo de pesquisa chamada de documental pode ser aplicada aos documentos que ainda não receberam um tratamento analítico. No caso dos *logs* eles transmitem exatamente o que foi pesquisado pelo usuário.

Para Godoy (1995), há diversos pontos a se observar em um documento que ainda não foi tratado. Fazendo com que ele sempre possa trazer novas descobertas e contribuir para o desenvolvimento das ciências.

Em relação ao objetivo, a pesquisa é do tipo descritiva, pois visa descrever a forma de linguagem utilizada pelos usuários e quais as diferenças entre estas e a linguagem de indexação.

O *corpus* foi constituído pelos *logs* correspondentes a cinco dias de pesquisa no portal LexML do Senado Federal Brasileiro, do dia 11 até o dia 17 de outubro de 2017.

O limite temporal se deve ao grande volume de informação disponível nos *logs* e ao tempo disponível para realização da pesquisa que, por se tratar de um trabalho de conclusão e curso, foi de aproximadamente seis meses. Destacamos, que o arquivo log de cada dia é pode conter de 800 a 1600 expressões de busca.

Os *logs* trazem especificamente o modo como o usuário fez a pesquisa, como visto anteriormente, eles são o rastro deixado pelos usuários no sistema. Assim sendo são uma fonte direta para identificação e coleta de termos por eles utilizados em suas buscas. Após o recebimento dos *logs*, em formato de texto e com todas as marcações do sistema, eles foram extraídos com a ajuda de um software extrator de logs criado, por um analista de sistemas, especialmente para realização da pesquisa, que organizou os dados em uma planilha do Excel, para facilitar a identificação das expressões de busca. Como na imagem 1 abaixo.

Imagem 1 – Exemplo de *logs*

IP	ANO	MÊS	DIA	HORA	MINUTO	SEGUNDO	PAÍS DE ORIGEM	EXPRESSÃO DE BUSCA
172.xx.y.114	2017	10	12	6	20	0	Brasil	"penha"
168.xx.yyy.19	2017	10	11	6	37	52	Brasil	"direto registral"
200.xxx.yy.260	2017	10	11	6	49	8	Brasil	"polícia penal"
187.xx.yyy.28	2017	10	11	6	56	11	Brasil	"HC 105.739"
187.xx.yyy.28	2017	10	11	7	11	56	Brasil	"HC 105.739"
187.xx.yyy.28	2017	10	11	7	12	18	Brasil	"HC 87926"
187.xx.yyy.28	2017	10	11	7	41	25	Brasil	"HC 87926"
187.xx.yyy.246	2017	10	17	10	55	36	Brasil	"servio tulio ruiz"
187.xx.yyy.246	2017	10	17	10	56	37	Brasil	"eugenio cuello calon"

Fonte: Dados da pesquisa

### 3.1 Organização e Limpeza dos dados

Foram acrescentadas mais três colunas ao arquivo do excel. Uma contendo Uma contendo a área do direito do termo, outra com a classificação do termo na cdd

e a ultima com observações feitas durante a classificação tal como: se o termo é o nome de um autor da área do direito na coluna de observações constará autor, como na imagem 2 abaixo.

Imagem 2 – Exemplo da Classificação feita com a CDD DIR

IP	ANO	MÊS	DIA	HORA	MINUTO	SEGUNDO	PAÍS DE ORIGEM	EXPRESSÃO DE BUSCA	ÁREA DO DIREITO	CLASSIFICAÇÃO (CDD Dir)	OBSERVAÇÃO
172.xx.y.114	2017	10	12	6	20	0	Brasil	"penha"	Direito Penal	341.5561	
168.xx.yyy.19	2017	10	11	6	37	52	Brasil	"direito registral"	Direito civil	342.12	
200.xxx.yy.260	2017	10	11	6	49	8	Brasil	"polícia penal"	Direito Penal	341.5436	
187.xx.yyy.28	2017	10	11	6	56	11	Brasil	"HC 105.739"	Direito Processual	341.4327	
187.xx.yyy.28	2017	10	11	7	11	56	Brasil	"HC 105.739"	Direito Processual	341.4327	
187.xx.yyy.28	2017	10	11	7	12	18	Brasil	"HC 87926"	Direito Processual	341.4327	
187.xx.yyy.28	2017	10	11	7	41	25	Brasil	"HC 87926"	Direito Processual	341.4327	
187.xx.yyy.246	2017	10	17	10	55	36	Brasil	"servio tulio ruiz"	direito penal	341.5	Autor
187.xx.yyy.246	2017	10	17	10	56	37	Brasil	"eugenio cuello calon"	direito penal	341.5	Autor

Fonte: Dados da pesquisa

O procedimento de classificação das variações foi feito utilizando a CDD DIR, que é uma extensão da classificação do direito. Cada expressão de busca do *log* foi classificada de acordo com a área do direito correspondente, porém apenas os referentes ao Direito Penal foram utilizados para análise em profundidade nesta pesquisa.

Realizamos uma comparação das expressões de busca contidas nos *logs* com os descritores existentes no Vocabulário básico controlado para a indexação da informação documental das bibliotecas ligadas ao Senado Federal Brasileiro.

Para realizar esta comparação todos os logs de busca do Direito Penal foram selecionados e colocados em um novo arquivo excel. Após foi acessado o site com o vocabulário controlado do Senado Federal (VCB). No site selecionou-se a opção VCB e marcou-se a palavra descritores, após foi feita uma busca com os termos de busca de cada log e acrescentado em uma nova coluna na tabela se consta ou não no VCB. Foi observado também se o termo era o preferido ou variante e isto foi acrescentado em outra coluna.

Através deste comparativo foi possível observar se a linguagem utilizada na indexação corresponde à linguagem utilizada pelo usuário durante a busca, ou se

elas não coincidem. Após os termos restantes, foram analisadas as variações denominativas de acordo com a classificação de Freixa (2002).

A limitação quanto a área do direito e ao tipo de variação também se deve ao tempo, já que não seria viável trabalhar com a variação conceitual e com todas as áreas do direito dentro do tempo estipulado. Limitou-se assim a pesquisa a variação denominativa na área do Direito Penal.

A utilização desta metodologia possibilitou o alcance dos objetivos propostos no trabalho dentro do tempo disponível. Sendo assim considerada adequada para o presente estudo. Os resultados da pesquisa apresentados no capítulo 4 estão divididos por tipo de variação denominada conforme a ocorrência. Este está dividido em alterações gráficas que como visto anteriormente consistem em uma alteração na grafia da palavra, reduções, que como vimos é uma redução no termo, as alterações léxicas de unidades poliléxicas que são uma alteração na base ou extensão do sintagma.

### 3.2 Análise dos dados

A partir dos procedimentos metodológicos já descritos no Capítulo 3 obtiveram-se 218 expressões de busca, destas foram encontrados 7 conjuntos de expressões de busca, que dizem respeito ao Direito Penal. Devido ao fato de que as variantes precisam ser analisadas em pares, apenas aquelas que possuíam pelo menos um termo variante em um outro *log* de busca foram utilizadas no presente trabalho. As leis são uma exceção, pois leis que possuem um nome popular, mesmo que com a ocorrência de apenas um dos termos foram analisadas.

Para preparar os dados para análise, as expressões foram reunidas de duas formas:

- a) Variação de estratégia de busca na jornada do usuário (expressões reunidas por IP de usuário)
- b) Variação terminológica independente do usuário (diferentes denominações de um determinado termo, mas expressas por usuários diversos)

Será possível verificar isto nos quadros do capítulo posterior, onde são apresentados os resultados da pesquisa. Em relação à análise terminológica buscou-se encontrar pares de variantes independentes de terem sido ou não

empregadas pelo mesmo usuário. No capítulo 4 descrevemos a análise de conjuntos de expressões de busca à luz da variação terminológica.

## 4 RESULTADOS DA PESQUISA

Este capítulo apresenta os resultados obtidos após a excussão dos processos metodológicos do capítulo 4. Ele está subdividido conforme a classificação da variação denominativa dos conjuntos analisados.

### 4.1 Alterações Gráficas

As alterações gráficas são um tipo de variação denominativa mencionada no capítulo 2.3. Nesta subdivisão serão inclusos os conjuntos que tenham pelo menos um par de variantes que estejam nesta categoria.

O primeiro caso observado nesta categoria é o da Lei N 12.737, de 30 de novembro de 2012, que ficou conhecida como Lei Carolina Dieckman, também é uma alteração gráfica do tipo termo e forma artificial. Sendo que o termo é o nome popular da lei, isto porque o número da lei foi decidido na hora de sua elaboração. Está lei foi uma alteração no Código Penal para crimes informáticos. No caso desta lei, nos dados da pesquisa, foi encontrado apenas o seu nome popular e nenhuma variante como é possível observar no quadro 8 abaixo.

Quadro 8 – Extrato do log com a expressão “Lei Carolina Dieckmann”

IP	Ano	Mês	Dia	Hora	Min	Seg	Expressão de busca	Qtde. ocorrências
200.xxx.yy.13	2017	10	16	12	16	32	"Lei Carolina Dieckmann"	1

Fonte: dados da pesquisa

Aqui é possível observar a importância do acréscimo do nome popular como remissiva no sistema, pois o usuário pode não ter conhecimento prévio do número da lei. O usuário nem sempre é um especialista na área do direito, o que pode ser o caso acima, devido a utilização do nome popular da lei e não de seu número.

Este caso também nos mostra a importância da garantia do usuário. Já que, como disseram Oliveira e Araujo (2012), o usuário precisa ser considerado sempre, já que ele será o que se beneficiará ou não da decisão tomada. Não acrescentar o nome popular como uma remissiva no sistema de recuperação da informação seria

colocar barreira para um usuário não familiarizado com a área dificultando assim o acesso a informação.

No portal LexMI, o usuário parece ter encontrado a informação que buscava já que não realizou uma outra busca. Porém ao analisar a ficha catalográfica das leis indexadas no portal percebemos que elas não constam como descritores, mas sim como autoridade. Incluindo nome popular e número da lei possibilitando assim que ambos sejam recuperados, porém apenas como autoridade.

Um outro caso a ser analisado é a ocorrência de uma alteração gráfica em “Termo circunstanciado policia militar” que também está presente como “TCO policia militar”. Sendo que o subtipo desta alteração é termo e sigla, devido ao fato de TCO ser uma sigla para termo circunstanciado. Como pode ser observado no quadro 9 abaixo.

Quadro 9 – Extrato do log com a expressão “TCO policia militar”

IP	Ano	Mês	Dia	Hora	Min	Seg	Expressão de busca	Qtde. ocorrências
187.xxx.yy.124	2017	10	13	8	47	42	"Tco policia militar"	1
187.xxx.yy.124	2017	10	13	8	47	58	"Termo circunstanciado policia militar"	1

Fonte: dados da pesquisa

Através do quadro 9 fica evidente a importância do acréscimo das siglas como remissivas no sistema, pois o usuário pode utilizá-las durante a busca. Como Lancaster (1979), traz os termos são parte importante dos sistemas de recuperação da informação, pois sem eles a recuperação da informação não seria possível. Sem o acréscimo da sigla no sistema, a informação não seria recuperada, mostrando assim como a recuperação da informação seria diretamente afetada pela ausência da remissiva com a sigla.

Fica evidente também a importância da garantia do usuário durante a elaboração da linguagem controlada que será utilizada para indexação, pois ela torna possível observar a relevância que as siglas tem para os usuários daquele sistema de recuperação da informação.

Os usuários que não são da área podem ter o conhecimento unicamente da sigla através da mídia e outros meios de comunicação. A sigla também pode ser de uso interno da área, neste caso a probabilidade de um leigo a conhecer seria

menor. De qualquer forma, este tipo de alteração gráfica é bem comum, o que mostra que os usuários do portal LExML utilizam com frequência siglas dos termos pelos quais estão procurando.

Neste caso é possível observar que o mesmo usuário pesquisa das duas maneiras, o que pode indicar que ele não obteve o que procurava durante a primeira busca na qual ele utilizou a abreviação. Como a busca por este termo ocorreu apenas uma vez, não é possível inferir se a busca pela sigla deste termo em específico é frequente, porém com certeza seria uma melhoria a inclusão de siglas no portal LexML.

O termo código penal apareceu com e sem acentuação (“código penal” e “codigo penal”). Essa é uma alteração gráfica do subtipo presença e ausência de acentuação. Este subtipo não foi comum nos logs, pois apareceu apenas uma vez dentre as 7 expressões de busca analisadas.

Quadro 10 – Extrato do log com a expressão “código penal”

IP	Ano	Mês	Dia	Hora	Min	Seg	Expressão de busca	Qtde. ocorrências
200.xxx.yy.205	2017	10	13	8	58	15	"CÓDIGO PENAL"	2
200.xxx.yy.188	2017	10	16	16	17	34	"codigo penal"	1

Fonte: dados da pesquisa

No corpus deste trabalho este subtipo foi o de menor ocorrência, sendo esta apenas uma ocorrência, como pode ser observado no quadro 10. Entretanto na Tese de Laipelt (2015), este subtipo de alteração gráfica ficou em terceiro lugar no número de ocorrências, no primeiro corpus, juntamente com o subtipo termo e abreviação. O subtipo presença e ausência de acentuação ocorreu em primeiro lugar no segundo corpus, junto com o subtipo termo e sigla.

Comparando com a tese de Laipelt (2015), na qual foram utilizados dois *corpus*, no primeiro corpus utilizado por ele teve duas ocorrências sendo isto uma ocorrência a mais do que no corpus do presente trabalho. Podemos assim inferir que é um subtipo de ocorrência comum, apesar de isto não ser aparente nos dados coletados para a realização deste trabalho. É possível que a baixa ocorrência desse subtipo se deva à área do direito escolhida para análise, que na tese de Laipelt (2015) foi Direito do Trabalho e Direito Previdenciário.



Também é possível observar que o sistema deve possibilitar a recuperação da informação sem acentuações. Pois sem isso, o usuário pode acabar não tendo acesso ao documento do qual necessita caso não perceba o que faltou na hora da busca. No caso do sistema possibilitar esta busca mesmo sem a acentuação também estaria contribuindo para poupar o tempo do usuário. O que é algo muito importante dentro da biblioteconomia já que é uma das suas cinco leis preconizadas por Ranganathan.

Além disso a língua portuguesa está em constante modificação. As regras de acentuação se modificam com o tempo. Então na hora da busca o usuário pode ter dúvida quanto a acentuação ou não da palavra. Este provavelmente não foi o caso de “codigo penal”, já que o mais provável é que tenha sido um erro na hora da digitação, porém a outra possibilidade não pode ser completamente descartada.

O PLS 38/2017 é uma alteração gráfica de projeto de lei 378/2017 sendo seu subtipo termo e forma artificial.

Quadro 11 – Extrato do log com a expressão “Projeto de lei 378/2017

IP	Ano	Mês	Dia	Hora	Min	Seg	Expressão de busca	Qtde. ocorrências
189.xx.yy.155	2017	10	16	19	7	56	"PLS 378/2017"	1
54.xxx.yyy.82	2017	10	16	20	57	18	"378/2017"	2
54.xxx.yyy.82	2017	10	16	20	57	29	"378"	2
54.xxx.yyy.82	2017	10	17	9	19	22	"PLS 378"	2
187.xxx.yy.203	2017	10	17	12	23	37	"projeto de lei 378/2017"	1

Fonte: dados da pesquisa

Mais uma vez a relevância das siglas serem remissivas é evidente, já que PLS é sigla de Projeto de Lei. Observa-se também, no quadro acima, que os termos utilizando PLS tiveram maior ocorrência do que o utilizando projeto de lei. Apesar de que entre “projeto de lei 978/2017” e “PLS 378/2017” o número de ocorrências foi o mesmo.

As Alterações gráficas do subtipo termo e sigla foram uma ocorrência comum inclusive comparando com o *corpus* da tese de Laipelt (2015) apresentando a segunda maior ocorrência, no primeiro corpus, e ficando em primeiro junto com o subtipo presença e ausência de acentuação, no segundo corpus.

No corpus deste trabalho houve duas ocorrências deste subtipo sendo o subtipo de maior ocorrência dentro das variações gráficas. Os outros subtipos tiveram apenas uma ocorrência, talvez devido a delimitação de dias utilizada para o presente estudo. Laipelt (2015) identificou que há uma tendência maior na utilização de siglas do que da forma completa, o que pode explicar o motivo de ter sido o subtipo de maior ocorrência em ambos os trabalhos.

## 4.2 Reduções

Nesta categoria foram analisados os conjuntos classificados como como reduções de acordo com o capítulo 2.3. Como é possível observar ocorreram diferentes tipos de variação em um mesmo termo como é o caso a seguir.

O primeiro caso a ser analisado nesta categoria é o da “Lei Maria da Penha” termo do qual ocorre uma outra variante dentro dos dados da pesquisa. Este termo seria “Maria da Penha” que é uma redução de base do subtipo [Substantivo + Substantivo] = [Substantivo 2]. O que ocorre também com o termo “Penha” que seria o mesmo subtipo de variação em relação ao termo “Maria da Penha”.

Quadro 12 – Extrato do log com a expressão “Lei Maria da Penha

IP	Ano	Mês	Dia	Hora	Min	Seg	Expressão de busca	Qtde. ocorrências
172.xx.y.114	2017	10	11	6	25	40	"penha"	5
172.xx.y.57	2017	10	11	8	55	19	"Lei maria da penha"	2
177.xx.yyy.125	2017	10	11	13	23	24	"maria da penha"	1
187.xx.yyy.56	2017	10	11	15	39	22	"violência domestica"	2

Fonte: dados da pesquisa

É possível observar no quadro 12 que o termo mais reduzido é o de maior ocorrência entre todos eles. Isto pode interferir na recuperação da informação, já que geralmente as reduções deixam o termo mais genérico. Isso, segundo Laipelt (2015), o uso de reduções acaba na maioria das vezes gerando perda de sentido do termo. Sendo assim para Laipelt (2015) a utilização das reduções deve ser considerada cuidadosamente para que não haja perda de sentido nas remissivas utilizadas.

Cabe mencionar que isto pode ser um problema na hora de recuperar a informação. Quando um termo é muito reduzido, ser muito genérico e recuperar

uma quantidade excessiva de informação, desnecessária ao usuário. Outra possibilidade é que o usuário não recupere a informação, pois estas reduções em geral não constam no sistema, nem mesmo como remissivas.

A Lei nº. 11.340, de 7 de agosto de 2006 ficou conhecida como “Lei Maria da Penha”, pois foi originada do caso de violência doméstica sofrido por Maria da Penha, que foi solucionado em 2002, ou seja, 4 anos antes da criação da lei.

Entretanto o nome popular da lei não é mencionado em momento algum na redação da lei. Nesse caso é possível observar a importância do princípio do uso, pois o número da lei não apareceu em momento algum no corpus da pesquisa.

Outro caso de redução é o do termo “projeto de Lei 378/2017”, que pode ser visto no quadro 11, é uma redução de estrutura do subtipo [substantivo 1 + substantivo 2] = [substantivo] como 378/ 2017 e apenas como 378, pois fica apenas a numeração.

Neste caso, com apenas a numeração, o termo fica muito genérico, e o usuário pode acabar recuperando coisas que não serão úteis a ele, como projetos e leis mais antigos. Pois algum projeto anterior pode ter sido numerado da mesma forma, caso seja um projeto municipal, por exemplo, o que o usuário procura.

Entretanto, novamente é possível se observar uma predominância das reduções nas buscas dos usuários, sendo, inclusive uma das mais frequentes, a utilização apenas do número do projeto lei. É possível que o usuário não seja um especialista na área e lembre do número do projeto e não do ano. Este projeto de lei está relacionado à lei do desarmamento, que também é conhecida como estatuto do desarmamento.

#### 4.3 Alterações léxicas de unidades poliléticas

É apresentada a análise de apenas um par de variantes, pois é o único par que ocorre na categoria alterações léxicas de unidades poliléticas. O par mencionado é o do termo “lei do desarmamento” como pode ser observado no quadro 13 abaixo

Quadro 13 – Extrato do log com a expressão “Lei do Desarmamento

IP	Ano	Mês	Dia	Hora	Min	Seg	Expressão de busca	Qtde. ocorrências
----	-----	-----	-----	------	-----	-----	--------------------	-------------------

168.xxx.yy.12	2017	10	18	0	18	53	"lei do desarmamento"	1
187.xx.yyy.207	2017	10	16	19	12	30	"estatuto desarmaamento"	1
200.xxx.yy.222	2017	10	17	18	47	11	"estatuto do desarmamento consulta"	1

Fonte: dados da pesquisa

Três usuários diferentes tentaram encontrar a mesma lei, como é possível observar no quadro 13. Porém, o primeiro utilizou “lei do desarmamento”, e o segundo “estatuto desarmaamento”. No segundo caso ele provavelmente buscava por estatuto do desarmamento, porém cometeu um erro de digitação.

“Lei do desarmamento” seria uma alteração léxica de unidades poliléticas do subtipo [substantivo + sintagma preposicional (prep. + art. + substantivo)] termo estatuto do desarmamento.

Este subtipo de alteração léxica de unidade polilética é incomum, já que em ambos os corpus utilizados por Laipelt (2015), ele não apareceu nenhuma vez, e no corpus do presente trabalho ele aparece apenas uma vez.

Porém isto não quer dizer que estes tipos de variantes não devam ser inclusas nas linguagens controladas. Mesmo que apareçam com menor frequência, elas ainda ocorrem. Como foi mencionado ao longo do trabalho, o usuário deve ser a prioridade e no caso de leis que tenham mais de um nome popular, como é o caso da lei 10.826, de 22 de dezembro de 2003, que também é conhecida como Lei do Desarmamento. Cabe destacar que ela é mais conhecida como estatuto do desarmamento, apesar de ser uma lei. Esta é a lei que trata do registro, posse e comercialização de armas de fogo e munição.

Também foi possível observar, neste caso, que os usuários utilizaram diferentes estratégias de busca. Cada um deles, apesar de procurarem pela mesma lei, realizou a busca de sua própria maneira. Um utilizando o termo Lei, o outro, estatuto e o ultimo utilizando o termo estatuto e acrescentando o termo consulta.

No ultimo caso, o usuário poderia estar pesquisando especificamente pela consulta pública sobre o estatuto. Ele também poderia querer apenas consultar o estatuto do desarmamento. Independente disto, é possível observar que ele utiliza dois termos distintos para tentar expressar sua necessidade de busca. Isto não será aprofundado, pois não é o objetivo do presente trabalho, porém achou-se importante mencionar isto na análise deste termo em específico.

#### 4.4 Casos excepcionais

Nesta categoria será analisado o caso dos termos armamento e desarmamento, que não se enquadram na classificação criada por Freixa (2002). Eles apareceram uma vez durante os dados da pesquisa, porém, foi possível observar que ambos os usuários estavam buscando pela mesma coisa, porém com termos semanticamente diferentes.

Quadro 14 – Extrato do log com relação de Oposição

IP	Ano	Mês	Dia	Hora	Min	Seg	Expressão de busca	Qtde. ocorrências
,187.xx.yy.255	2017	10	17	13	54	52	"armamento"	3
189.xx.yy.86	2017	10	17	17	13	50	"desarmamento"	1

Fonte: dados da pesquisa

Como é possível observar através do quadro acima houve uma predominância da procura por armamento. Observa-se três ocorrências deste termo e apenas uma do termo desarmamento. Neste caso, os termos estão semanticamente relacionados. Esta relação entre os termos, conhecida como relação de oposição, não é prevista na classificação de variações denominativas de Freixa (2002). De acordo com Gomes (1990), dentro dos relacionamentos lógicos no Tesauro se encontram o relacionamento genérico/específico, relacionamento analítico e o relacionamento de oposição. De acordo com Khoo e Jin-Cheon (2006), a relação de oposição é também conhecida como antonímia e ela é a mais estudada. Também de acordo com os autores esse tipo de relação é mais facilmente compreendido pelas pessoas, pois consiste em analisar pares que são opostos um ao outro.

Segundo Laipelt (2015), as relações que não apresentam hierarquia entre si nos tesauros são chamadas de relações associativas. De acordo com a IFLA (2012) as relações associativas mais comuns são Causa/Efeito, Processo/Agente, Ação/Produção da ação, Ação/ Paciente ou Objetivo, Conceito ou coisa/ Propriedades, Ação ou Coisa/ Contra Agente, Coisa/ Suas Partes, Matéria Prima/ Produto, Ação/Propriedade, Campo de estudo/ Objeto ou Fenômenos estudados.

A inclusão de termos opostos do mesmo campo semântico em uma linguagem controlada, e conseqüentemente na indexação de obras, possibilita que a informação seja recuperada pelos usuários.

## 5 CONSIDERAÇÕES FINAIS

Apesar da necessidade de se buscar novas formas de acompanhar a forma com que os usuários fazem sua pesquisa, ainda não é frequente o uso de logs para isso. Apesar da contribuição que os logs podem trazer para a indexação em sistemas de recuperação da informação, ainda se tem poucos estudos utilizando logs como fonte de coleta de dados.

Embora seja possível perceber a importância da guarda dos logs para estudos que possibilitam melhorias no sistema, e até mesmo o estudo sobre o acesso a artigos de periódicos, ainda não é comum a guarda desses. Isso ocorre porque os logs ocupam espaço no servidor, porém com parâmetros para a guarda dos mesmos, que defina por quanto tempo isso ocorrerá, esse problema seria evitado e possibilitaria um maior número de estudos utilizando-os.

Em sua tese Laipelt (2015), utilizou logs para analisar a variação denominativa na área do Direito do Trabalho e do Direito Previdenciário. Apesar de serem áreas diferentes do direito, pode-se notar algumas similaridades entre os resultados obtidos por Laipelt (2015) e os resultados obtidos no presente trabalho.

Quanto aos estudos de variação na área jurídica, o mais atual encontrado pela autora do presente trabalho foi a dissertação de Luciana Monteiro Krebs, que estudou a variação conceitual dos termos do domínio jurídico.

Através do presente trabalho ficou evidente a importância da garantia do usuário na elaboração de vocabulários controlados, ou mesmo para a gestão de catálogos de autoridades. Também ficou claro que os logs são de extrema importância neste sentido, pois com as modificações que ocorreram na relação entre o bibliotecário e o usuário, os logs se tornaram uma maneira muito eficaz de se verificar a terminologia por eles utilizada.

Os logs possibilitam que o usuário seja considerado dentro de uma linguagem padronizada. Sendo assim evita-se perda na hora de recuperar a informação, pois o bibliotecário e o usuário utilizam a mesma linguagem.

É possível perceber que a Terminologia tem grande influência nos sistemas de recuperação da informação. Podendo contribuir com melhorias no sistema de recuperação da informação através do acréscimo de remissivas. Cabe destacar que a variação denominativa influencia diretamente na qualidade da recuperação da informação, podendo inclusive determinar a recuperação ou não da informação.

Os logs podem contribuir na escolha de descritores e também apresentarem novos termos que podem vir a fazer parte da rede de remissivas. Foi possível perceber neste estudo que os usuário da área do Direito Penal do portal LexML, em sua maioria, utilizam, para a busca de informação, as siglas. Na pesquisa de Laipelt (2015), o uso de siglas pelos usuários do portal LexML das áreas do Direito do Trabalho e do Direito Previdenciário também se mostraram frequentes. Por isso, é interessante inserir as siglas nos vocabulários controlados e/ou sistemas de remissivas, sempre que forem identificadas, para facilitar a recuperação da informação pelos usuários.

Nos resultados do presente trabalho também foi constatada a presença de reduções de extensão e reduções de base e uma alteração léxica de unidade poliléxica. Quanto a pesquisa no vocabulário controlado do Senado Federal 3 dos dezesseis termos dos sete conjuntos foram encontrados como autoridade, 3 como termo preferido, 1 como variante e 9 não foram encontrados.

Isto demonstra o impacto direto da variação denominativa na recuperação da informação, pois os 9 termos que não foram encontrados indicam que o usuário precisará fazer uma nova pesquisa utilizando outro termo, ou que o usuário não encontrará a informação.

Outro aspecto a destacar é que nem todas as variantes identificadas neste trabalho se enquadraram na classificação de Freixa (2002). Por isso utilizamos a sistematização de tipologias de relações semânticas associativas da IFLA (2012) para análise dos termos não contemplados pela classificação de Freixa (2002). Nesta pesquisa, apenas a relação de oposição entre termos não pode ser enquadrada na classificação de Freixa (2002).

É importante mencionar que, sem o controle de vocabulário, a recuperação da informação se torna difícil, algumas vezes quase impossível. Por isso, cabe aos bibliotecários realizarem o planejamento e controle de termos e descritores utilizados para indexação de obras nas instituições, seja a partir da elaboração de tesauros propriamente ditos ou gestão de catálogos de autoridades. Como cada biblioteca é diferente, é necessário que a linguagem seja pensada de acordo com as necessidades de seus usuários.

Futuros trabalhos podem vir a ser feitos utilizando logs. Estes podem vir a estudar as variantes de outras áreas do conhecimento tais como dermatologia, matemática, entre outras. Também podem vir a estudar outras áreas do Direito

como o Direito Civil. Também é possível realizar estudos sobre as estratégias de busca realizadas pelos usuários, contribuindo assim para a constante melhorias nos sistema de recuperação da informação. Além de estudos que quantifiquem os autores mais procurados em determinada área do conhecimento e em um sistema de recuperação da informação específico.



## REFERÊNCIAS

ARAÚJO, Vera Maria Araujo Pigozzi. Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos. **Transinformação**, v. 24, n. 2, 2012. Disponível em: <<http://www.brapci.inf.br/v/a/20379>>. Acesso em: 15 out. 2017.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676: Métodos para análise de documentos**: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro. 1992.

CABRÉ, Maria Teresa. El principio de poliedricidad: la articulación de lo discursivo, lo cognitivo y lo lingüístico en Terminología (I). **Ibérica**, v.1, n. 16, 2008.

CABRÉ, Maria Teresa. LA TEORÍA COMUNICATIVA DE LA TERMINOLOGÍA, UNA APROXIMACIÓN LINGÜÍSTICA A LOS TÉRMINOS. **Revue française de linguistique appliquée**, v.14, n.2, 2009, p. 9-15.

CAFE, Lígia Maria Arruda. A interoperabilidade semântica entre acervos de museus: discutindo o caso dos museus da imagem e do som. **Em Questão**, v. 23, n. 1, 2017, p. 113 - 128. Disponível em: <<http://www.brapci.inf.br/v/a/21144>>. Acesso em: 15 out 2017.

CARNEIRO, Marília Vidigal. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, v. 14, n. 2, p. 221-241, 1985. Disponível em: <<http://www.brapci.inf.br/v/a/2649>>. Acesso em: 11 nov. 2017.

FERRO, Vinicius. **A indexação e o usuário**: análise de expressões de busca do direito penal no portal LEXML. 2014. 124 f. Trabalho de conclusão de curso (bacharelado) - Faculdade de biblioteconomia e comunicação, Porto Alegre: 2014.

FREIXA, Judit. Variación denominativa y biunivocidad en el lenguaje de las matemáticas. **Revista española de lingüística**, v. 36, 2006, 189-215.

FUGITA, Mariângela Spotti Lopes. A política de indexação para representação e recuperação da informação. IN: LEIVA, Isidoro Gil; FUJITA, Mariângela Spotti Lopes. **Política de Indexação**. Cultura acadêmica: Marília, 2012. Cap. 1, p. 17-31.

FUGITA, Mariângela Spotti Lopes. Organização e representação do conhecimento no brasil: análise de aspectos conceituais e da produção científica do enancib no período de 2005 a 2007. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 1, n. 1, p. 1-32, 2008. Disponível em: <<http://www.brapci.inf.br/v/a/7781>>. Acesso em: 05 Jul 2018.

GABRIEL JUNIOR, Rene Faustino.; LAIPELT, Rita do Carmo Ferreira. Thesa: ferramenta para construção de tesauro semântico aplicado interoperável. **Revista P2P e INOVAÇÃO**, v. 3, n. 2., 2017, p. 124-145. Disponível em: <<http://www.brapci.inf.br/v/a/22911>>. Acesso em: 11 nov. 2017.

GABRIEL JUNIOR, Rene Faustino. Aproximação da bibliometria e recuperação de informação na brapci. In: **Encontro Brasileiro de Bibliometria e Cientometria**, 4, 2014, Recife. **Anais...** Recife: UFEP, 2014.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (Org.). **Métodos de pesquisa**. Porto Alegre: Ed. UFRGS, 2009. Disponível em: <[www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf](http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf)>. Acesso em: 10 Nov. 2017.

GODOY, Arilda Schmidt. Pesquisa qualitativa: tipos fundamentais. *Revista de Administração de Empresas*, São Paulo, v. 35, n. 3, p. 20-29, maio/jun. 1995.

GUEDES, Roger de Miranda; DIAS, Eduardo José Wense. Indexação social: abordagem conceitual. *Revista ABC*, Florianópolis, v. 15, n. 1, p. 39-53, 2010. Disponível em: <[http://www.brapci.inf.br/\\_repositorio/2010/06/pdf\\_fcb17df2cd\\_0010808.pdf](http://www.brapci.inf.br/_repositorio/2010/06/pdf_fcb17df2cd_0010808.pdf)>. Acesso em: 10 nov. 2017

GUIM, Vera Lucia Ribeiro; FUJITA, Mariângela Spotti Lopes. Bibliotecas escolares e as linguagens de indexação. **Biblioteca Escolar em Revista**, v. 4, n. 2, 2016. Disponível em: <<http://www.brapci.inf.br/v/a/21367>>. Acesso em: 13 out. 2017.

International Federation of Library Associations and Institutions. **Requisitos Funcionales Para Datos de Autoridad de Materia (FRASAD): un modelo conceptual**. Netherlands: IFLA, 2012, 66 p.

KURAMOTO, Hélio. Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais. **Ciência da Informação**, v.25, n. 2, 1995.

LAIPELT, Rita do Carmo Ferreira. A análise de logs como estratégia para a realização da garantia do usuário. **Em Questão**, v. 21, n. 3, p. 150-170. . Disponível em: <<http://www.brapci.inf.br/v/a/18673>>. Acesso em: 11 Nov. 2017.

LAIPELT, Rita do Carmo Ferreira. Metodologia para seleção de termos equivalente e descritores de tesauro: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário. 2015. 213 f. Tese (Doutorado em Linguística Aplicada) - Programa

de Pós Graduação em Linguística Aplicada, UNISINOS, São Leopoldo, 2015. Disponível em: <[http://www.repositorio.jesuita.org.br/bitstream/handle/UNISINOS/4853/RITA+DO+CARMO+FERREIRA+LAIPILT\\_.pdf;jsessionid=F2524050225B3D7776DED10021519629?sequence=1](http://www.repositorio.jesuita.org.br/bitstream/handle/UNISINOS/4853/RITA+DO+CARMO+FERREIRA+LAIPILT_.pdf;jsessionid=F2524050225B3D7776DED10021519629?sequence=1)>. Acesso em 30 set. 2017.

LANCASTER, Frederick Wilfrid. **Indexação e Resumo**: teoria e prática. Brasília: Briquet de Lemos, 2004.

LANCASTER, Frederick Wilfrid. **Information Retrieval Systems**: characteristics, testing and evaluation. New York: J. Wiley, 1979. 2nd ed. 381 p.

Manual de elaboração de tesouros monolíngües. GOMES. Hagar Espanha, coord. **Manual de elaboração de tesouros monolíngües**. Brasília: PNB, 1990, 78p.

Moreira, Manoel Palhares; Moura, Maria Aparecida. Construindo tesouros a partir de tesouros existentes: a experiência do TCI - Tesouro em Ciência da Informação. **DataGramaZero**, v.7, n.4, 2006.

NICHOLAS, David; HUNTIGTON, Paul; WATKINSON. Scholarly journal usage: the results of deep log analysis. *Journal of documentation*, London, V. 61, n.2, p. 248-280, 2005.

OLIVEIRA, Dalgiza Andrade; ARAUJO, Ronaldo Ferreira de. construção de linguagens documentárias em sistemas de recuperação da informação: a importância da garantia do usuário. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. 34, p.17-30.

PASSOS, Paula Caroline Schifino Jardim.; CAREGNATO, Sônia Elisa. Análise dos sistemas de busca de revistas científicas eletrônicas. **Revista Digital de Biblioteconomia & Ciência da Informação**, v. 16, n. 2. Disponível em: <<http://www.brapci.inf.br/v/a/29881>>. Acesso em: 14 Maio 2018.